

SEQUENCE LISTING

| CERTIFICATE OF MAILING | |
|---|--|
| I hereby certify that this correspondence is being deposited with the United States Postal Service as First Class Mail in an envelope addressed to: Commissioner of Patents and Trademarks. | |
| Washington, DC 20231 on | 1/8/93 |
| LIMBACH & LIMBACH | |
| Dated: 1/8/93 | By: <i>[Signature]</i> Name: <i>[Signature]</i> |

- (1) GENERAL INFORMATION:
 - (i) APPLICANT: Donson, Jon
Dawson, William O.
Grantham, George L.
Turpen, Thomas H.
Turpen, Ann Myers
Garger, Stephen J.
Grill, Laurence K.
 - (ii) TITLE OF INVENTION: RECOMBINANT PLANT VIRAL NUCLEIC ACIDS
 - (iii) NUMBER OF SEQUENCES: 11
 - (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Limbach & Limbach
 - (B) STREET: 2001 Ferry Building
 - (C) CITY: San Francisco
 - (D) STATE: CAL
 - (F) ZIP: 94111
 - (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: Patent in Release #1.0, Version #1.25
 - (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER:
 - (B) FILING DATE:
 - (C) CLASSIFICATION:
 - (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 600,244
 - (B) FILING DATE: 22-OCT-1990
 - (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 641,617
 - (B) FILING DATE: 16-JAN-1991
 - (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 310,881
 - (B) FILING DATE: 17-FEB-1989
 - (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 160,766
 - (B) FILING DATE: 26-FEB-1988
 - (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 160,771
 - (B) FILING DATE: 26-FEB-1988

- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 347,637
 - (B) FILING DATE: 05-MAY-1989
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 363,138
 - (B) FILING DATE: 08-JUN-1989
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 219,279
 - (B) FILING DATE: 15-JUL-1988
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Halluin, Albert P.
 - (B) REGISTRATION NUMBER: 28,957
 - (C) REFERENCE/DOCKET NUMBER: BIOG-20121 USA
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: 415-433-4150
 - (B) TELEFAX: 415-433-8716
- (2) INFORMATION FOR SEQ ID NO: 1:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 4 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: peptide
 - (iii) HYPOTHETICAL: NO
 - (iv) ANTI-SENSE: NO
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:
Pro Xaa Gly Pro
1
- (2) INFORMATION FOR SEQ ID NO: 2:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 13 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (iii) HYPOTHETICAL: NO
 - (iv) ANTI-SENSE: NO
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 886 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: Chinese cucumber

(vii) IMMEDIATE SOURCE:
 (B) CLONE: alpha-trichosanthin

(ix) FEATURE:
 (A) NAME/KEY: CDS (B) LOCATION: 8. .877
 (B) LOCATION: 8. .877

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

CTCGAGG ATG ATC AGA TTC TTA GTC CTC TCT TTG CTA ATT CTC ACC CTG9
 Met Ile Arg Phe Leu Val Leu Ser Leu Leu Ile Leu Thr Leu
 1 5 10

TTC CTA ACA ACT CCT GCT GTG GAG GGC GAT GTT AGC TTC CGT TTA TC97
 Phe Leu Thr Thr Pro Ala Val Glu Gly Asp Val Ser Phe Arg Leu Ser
 15 20 25 30

GGT GCA ACA AGC AGT TCC TAT GGA GTT TTC ATT TCA AAT CTG AGA AA45
 Gly Ala Thr Ser Ser Ser Tyr Gly Val Phe Ile Ser Asn Leu Arg Lys
 35 40 45

GCT CTT CCA AAT GAA AGG AAA CTG TAC GAT ATC CCT CTG TTA CGT TC93
 Ala Leu Pro Asn Glu Arg Lys Leu Tyr Asp Ile Pro Leu Leu Arg Ser
 50 55 60

TCT CTT CCA GGT TCT CAA CGC TAC GCA TTG ATC CAT CTC ACA AAT TA41
 Ser Leu Pro Gly Ser Gln Arg Tyr Ala Leu Ile His Leu Thr Asn Tyr
 65 70 75

GCC GAT GAA ACC ATT TCA GTG GCC ATA GAC GTA ACG AAC GTC TAT AT289

Ala Asp Glu Thr Ile Ser Val Ala Ile Asp Val Thr Asn Val Tyr Ile
 80 85 90
 ATG GGA TAT CGC GCT GGC GAT ACA TCC TAT TTT TTC AAC GAG GCT TCB7
 Met Gly Tyr Arg Ala Gly Asp Thr Ser Tyr Phe Phe Asn Glu Ala Ser
 95 100 105 110
 GCA ACA GAA GCT GCA AAA TAT GTA TTC AAA GAC GCT ATG CGA AAA GBB5
 Ala Thr Glu Ala Ala Lys Tyr Val Phe Lys Asp Ala Met Arg Lys Val
 115 120 125
 ACG CTT CCA TAT TCT GGC AAT TAC GAA AGG CTT CAA ACT GCT GCG GCB3
 Thr Leu Pro Tyr Ser Gly Asn Tyr Glu Arg Leu Gln Thr Ala Ala Gly
 130 135 140
 AAA ATA AGG GAA AAT ATT CCG CTT GGA CTC CCA GCT TTG GAC AGT GCB1
 Lys Ile Arg Glu Asn Ile Pro Leu Gly Leu Pro Ala Leu Asp Ser Ala
 145 150 155
 ATT ACC ACT TTG TTT TAC TAC AAC GCC AAT TCT GCT GCG TCG GCA CCB9
 Ile Thr Thr Leu Phe Tyr Tyr Asn Ala Asn Ser Ala Ala Ser Ala Leu
 160 165 170
 ATG GTA CTC ATT CAG TCG ACG TCT GAG GCT GCG AGG TAT AAA TTT ABB7
 Met Val Leu Ile Gln Ser Thr Ser Glu Ala Ala Arg Tyr Lys Phe Ile
 175 180 185 190
 GAG CAA CAA ATT GGG AAG CGC GTT GAC AAA ACC TTC CTA CCA AGT TBB5
 Glu Gln Gln Ile Gly Lys Arg Val Asp Lys Thr Phe Leu Pro Ser Leu
 195 200 205
 GCA ATT ATA AGT TTG GAA AAT AGT TGG TCT GCT CTC TCC AAG CAA ABB3
 Ala Ile Ile Ser Leu Glu Asn Ser Trp Ser Ala Leu Ser Lys Gln Ile
 210 215 220
 CAG ATA GCG AGT ACT AAT AAT GGA CAG TTT GAA ACT CCT GTT GTG CBB1
 Gln Ile Ala Ser Thr Asn Asn Gly Gln Phe Glu Thr Pro Val Val Leu
 225 230 235
 ATA AAT GCT CAA AAC CAA CGA GTC ATG ATA ACC AAT GTT GAT GCT GCB9
 Ile Asn Ala Gln Asn Gln Arg Val Met Ile Thr Asn Val Asp Ala Gly
 240 245 250
 GTT GTA ACC TCC AAC ATC GCG TTG CTG CTG AAT CGA AAC AAT ATG GCB7

Val Val Thr Ser Asn Ile Ala Leu Leu Leu Asn Arg Asn Asn Met Ala
 255 260 265 270
 GCC ATG GAT GAC GAT GTT CCT ATG ACA CAG AGC TTT GGA TGT GGA AGG
 Ala Met Asp Asp Asp Val Pro Met Thr Gln Ser Phe Gly Cys Gly Ser
 275 280 285
 TAT GCT ATT TAGTAACTCG AG 886
 Tyr Ala Ile
 290

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 289 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Ile Arg Phe Leu Val Leu Ser Leu Leu Ile Leu Thr Leu Phe Leu
 1 5 10 15
 Thr Thr Pro Ala Val Glu Gly Asp Val Ser Phe Arg Leu Ser Gly Ala
 20 25 30
 Thr Ser Ser Ser Tyr Gly Val Phe Ile Ser Asn Leu Arg Lys Ala Leu
 35 40 45
 Pro Asn Glu Arg Lys Leu Tyr Asp Ile Pro Leu Leu Arg Ser Ser Leu
 50 55 60
 Pro Gly Ser Gln Arg Tyr Ala Leu Ile His Leu Thr Asn Tyr Ala Asp
 65 70 75 80
 Glu Thr Ile Ser Val Ala Ile Asp Val Thr Asn Val Tyr Ile Met Gly
 85 90 95
 Tyr Arg Ala Gly Asp Thr Ser Tyr Phe Phe Asn Glu Ala Ser Ala Thr
 100 105 110
 Glu Ala Ala Lys Tyr Val Phe Lys Asp Ala Met Arg Lys Val Thr Leu
 115 120 125
 Pro Tyr Ser Gly Asn Tyr Glu Arg Leu Gln Thr Ala Ala Gly Lys Ile
 130 135 140
 Arg Glu Asn Ile Pro Leu Gly Leu Pro Ala Leu Asp Ser Ala Ile Thr
 145 150 155 160

| | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Thr | Leu | Phe | Tyr | Tyr | Asn | Ala | Asn | Ser | Ala | Ala | Ser | Ala | Leu | Met | Val |
| | | | | | 165 | | | | 170 | | | | | 175 | |
| Leu | Ile | Gln | Ser | Thr | Ser | Glu | Ala | Ala | Arg | Tyr | Lys | Phe | Ile | Glu | Gln |
| | | | 180 | | | | | 185 | | | | | 190 | | |
| Gln | Ile | Gly | Lys | Arg | Val | Asp | Lys | Thr | Phe | Leu | Pro | Ser | Leu | Ala | Ile |
| | | 195 | | | | | 200 | | | | | 205 | | | |
| Ile | Ser | Leu | Glu | Asn | Ser | Trp | Ser | Ala | Leu | Ser | Lys | Gln | Ile | Gln | Ile |
| | 210 | | | | | 215 | | | | | 220 | | | | |
| Ala | Ser | Thr | Asn | Asn | Gly | Gln | Phe | Glu | Thr | Pro | Val | Val | Leu | Ile | Asn |
| 225 | | | | | 230 | | | | | 235 | | | | | 240 |
| Ala | Gln | Asn | Gln | Arg | Val | Met | Ile | Thr | Asn | Val | Asp | Ala | Gly | Val | Val |
| | | | | 245 | | | | | 250 | | | | | 255 | |
| Thr | Ser | Asn | Ile | Ala | Leu | Leu | Leu | Asn | Arg | Asn | Asn | Met | Ala | Ala | Met |
| | | | 260 | | | | | | 265 | | | | 270 | | |
| Asp | Asp | Asp | Val | Pro | Met | Thr | Gln | Ser | Phe | Gly | Cys | Gly | Ser | Tyr | Ala |
| | | 275 | | | | | 280 | | | | | 285 | | | |

Ile

- (2) INFORMATION FOR SEQ ID NO: 5:
- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1452 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
 (A) ORGANISM: *Oryza sativa*
- (vii) IMMEDIATE SOURCE:
 (B) CLONE: alpha-amylase
- (ix) FEATURE:
 (A) NAME/KEY: CDS (B) LOCATION: 12. .1316
 (B) LOCATION: 12. .1316
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:
- CCTCGAGGTG C ATG CAG GTG CTG AAC ACC ATG GTG AAC A CAC TTC TTG 50
- Met Gln Val Leu Asn Thr Met Val Asn Lys His Phe Leu

1

5

10

TCC CTT TCG GTC CTC ATC GTC CTC CTT GGC CTC TCC TCC AAC TTG AC~~2~~8

Ser Leu Ser Val Leu Ile Val Leu Leu Gly Leu Ser Ser Asn Leu Thr
15 20 25

GCC GGG CAA GTC CTG TTT CAG GGA TTC AAC TGG GAG TCG TGG AAG GA~~3~~6

Ala Gly Gln Val Leu Phe Gln Gly Phe Asn Trp Glu Ser Trp Lys Glu
30 35 40 45

AAT GGC GGG TGG TAC AAC TTC CTG ATG GGC AAG GTG GAC GAC ATC GC~~4~~4

Asn Gly Gly Trp Tyr Asn Phe Leu Met Gly Lys Val Asp Asp Ile Ala
50 55 60

GCA GCC GGC ATC ACC CAC GTC TGG CTC CCT CCG CCG TCT CAC TCT GT~~2~~2

Ala Ala Gly Ile Thr His Val Trp Leu Pro Pro Pro Ser His Ser Val
65 70 75

GGC GAG CAA GGC TAC ATG CCT GGG CGG CTG TAC GAT CTG GAC GCG TC~~2~~0

Gly Glu Gln Gly Tyr Met Pro Gly Arg Leu Tyr Asp Leu Asp Ala Ser
80 85 90

AAG TAC GGC AAC GAG GCG CAG CTC AAG TCG CTG ATC GAG GCG TTC C~~2~~8

Lys Tyr Gly Asn Glu Ala Gln Leu Lys Ser Leu Ile Glu Ala Phe His
95 100 105

GGC AAG GGC GTC CAG GTG ATC GCC GAC ATC GTC ATC AAC CAC CGC AC~~3~~6

Gly Lys Gly Val Gln Val Ile Ala Asp Ile Val Ile Asn His Arg Thr
110 115 120 125

GCG GAG CAC AAG GAC GGC CGC GGC ATC TAC TGC CTC TTC GAG GGC G~~3~~4

Ala Glu His Lys Asp Gly Arg Gly Ile Tyr Cys Leu Phe Glu Gly Gly
130 135 140

ACG CCC GAC TCC CGC CTC GAC TGG GGC CCG CAC ATG ATC TGC CGC GA~~2~~2

Thr Pro Asp Ser Arg Leu Asp Trp Gly Pro His Met Ile Cys Arg Asp
145 150 155

GAC CCC TAC GGC CAT GGC ACC GGC AAC CCG GAC ACC GGC GCC GAC TC~~3~~0

Asp Pro Tyr Gly Asp Gly Thr Gly Asn Pro Asp Thr Gly Ala Asp Phe
160 165 170

GCC GCC GCG CCG GAC ATC GAC CAC CTC AAC AAG CGC GTC CAG CGG GA~~7~~8

Ala Ala Ala Pro Asp Ile Asp His Leu Asn Lys Arg Val Gln Arg Glu
175 180 185

CTC ATT GGC TGG CTC GAC TGG CTC AAG ATG GAC ATC GGC TTC GAC GGC 26
 Leu Ile Gly Trp Leu Asp Trp Leu Lys Met Asp Ile Gly Phe Asp Ala
 190 195 200 205
 TGG CGC CTC GAC TTC GCC AAG GGC TAC TCC GCC GAC ATG GCA AAC ACC 74
 Trp Arg Leu Asp Phe Ala Lys Gly Tyr Ser Ala Asp Met Ala Lys Ile
 210 215 220
 TAC ATC GAC GCC ACC GAG CCG AGC TTC GCC GTG CCC GAG ATA TCG ACC 22
 Tyr Ile Asp Ala Thr Glu Pro Ser Phe Ala Val Ala Glu Ile Trp Thr
 225 230 235
 TCC ATG GCG AAC GGC GGG GAC GGC AAG CCG AAC TAC GAC CAG AAC GCC 70
 Ser Met Ala Asn Gly Gly Asp Gly Lys Pro Asn Tyr Asp Gln Asn Ala
 240 245 250
 CAC CGG CAG GAG CTG GTC AAC TGG GTC GAT CGT GTC GGC GGC GCC ACC 18
 His Arg Gln Glu Leu Val Asn Trp Val Asp Arg Val Gly Gly Ala Asn
 255 260 265
 ACC AAC GGC ACG GCG TTC GAC TTC ACC ACC AAG GGC ATC CTC AAC GGC 6
 Ser Asn Gly Thr Ala Phe Asp Phe Thr Thr Lys Gly Ile Leu Asn Val
 270 275 280 285
 GCC GTG GAG GGC GAG CTG TGG CGC CTC CGC GGC GAG GAC GGC AAG GCC 14
 Ala Val Glu Gly Glu Leu Trp Arg Leu Arg Gly Glu Asp Gly Lys Ala
 290 295 300
 CCC GGC ATG ATC GGG TGC TGG CCG GCC AAG GCG ACG ACC TTC GTC GAG 62
 Pro Gly Met Ile Gly Trp Trp Pro Ala Lys Ala Thr Thr Phe Val Asp
 305 310 315
 AAC CAC GAC ACC GGC TCG ACG CAG CAC CTG TGG CCG TTC CCC TCC GAC 10
 Asn His Asp Thr Gly Ser Thr Gln His Leu Trp Pro Phe Pro Ser Asp
 320 325 330
 AAG GTC ATG CAG GGC TAC GCA TAC ATC CTC ACC CAC CCC GGC AAC ACC 58
 Lys Val Met Gln Gly Tyr Ala Tyr Ile Leu Thr His Pro Gly Asn Pro
 335 340 345
 TGC ATC TTG TAC GAC CAT TTC TTC GAT TGG GGT CTC AAG GAG GAG ACC 6
 Cys Ile Phe Tyr Asp His Phe Phe Asp Trp Gly Leu Lys Glu Glu Ile
 350 355 360 365

GAG CGC CTG GTG TCA ATC AGA AAC CGG CAG GGG ATC CAC CCG GCG AGC 34
 Glu Arg Leu Val Ser Ile Arg Asn Arg Gln Gly Ile His Pro Ala Ser
 370 375 380

GAG CTG CGC ATC ATG GAA GCT GAC AGC GAT CTC TAC CTC GCG GAG ATT 2
 Glu Leu Arg Ile Met Glu Ala Asp Ser Asp Leu Tyr Leu Ala Glu Ile
 385 390 395

GAT GGC AAG GTG ATC ACA AAG ATT GGA CCA AGA TAC GAC GTC GAA CTT 30
 Asp Gly Lys Val Ile Thr Lys Ile Gly Pro Arg Tyr Asp Val Glu His
 400 405 410

CTC ATC CCC GAA GGC TTC CAG GTC GTC GCG CAC GGT GAT GGC TAC GCT 8
 Leu Ile Pro Glu Gly Phe Gln Val Val Ala His Gly Asp Gly Tyr Ala
 415 420 425

ATC TGG GAG AAA ATC TGAGCGCACG ATGACGAGAC TCTCAGTTTTC 33
 Ile Trp Glu Lys Ile
 430 435

CCTGCGATTT TTACCCTGAC CGGTATACGT ATATACGTGC CGGCAACGAG
 CTGTATCCGA 1413

TCCGAATTAC GGATGCAATT GTCCACGAAG TCCTCGAGG 1452

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 434 amino acids
 (B) TYPE: amino acid
 (D) Topology: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

Met Gln Val Leu Asn Thr Met Val Asn Lys His Phe Leu Ser Leu Ser
 1 5 10 15

Val Leu Ile Val Leu Leu Gly Leu Ser Ser Asn Leu Thr Ala Gly Gln
 20 25 30

Val Leu Phe Gln Gly Phe Asn Trp Glu Ser Trp Lys Glu Asn Gly Gly
 35 40 45

Trp Tyr Asn Phe Leu Met Gly Lys Val Asp Asp Ile Ala Ala Ala Gly
 50 55 60

| | | | | | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Ile | Thr | His | Val | Trp | Leu | Pro | Pro | Pro | Ser | His | Ser | Val | Gly | Glu | Gln | 65 | 70 | 75 | 80 |
| Gly | Tyr | Met | Pro | Gly | Arg | Leu | Tyr | Asp | Leu | Asp | Ala | Ser | Lys | Tyr | Gly | 85 | 90 | 95 | |
| Asn | Glu | Ala | Gln | Leu | Lys | Ser | Leu | Ile | Glu | Ala | Phe | His | Gly | Lys | Gly | 100 | 105 | 110 | |
| Val | Gln | Val | Ile | Ala | Asp | Ile | Val | Ile | Asn | His | Arg | Thr | Ala | Glu | His | 115 | 120 | 125 | |
| Lys | Asp | Gly | Arg | Gly | Ile | Tyr | Cys | Leu | Phe | Glu | Gly | Gly | Thr | Pro | Asp | 130 | 135 | 140 | |
| Ser | Arg | Leu | Asp | Trp | Gly | Pro | His | Met | Ile | Cys | Arg | Asp | Asp | Pro | Tyr | 145 | 150 | 155 | 160 |
| Gly | Asp | Gly | Thr | Gly | Asn | Pro | Asp | Thr | Gly | Ala | Asp | Phe | Ala | Ala | Ala | 165 | 170 | 175 | |
| Pro | Asp | Ile | Asp | His | Leu | Asn | Lys | Arg | Val | Gln | Arg | Glu | Leu | Ile | Gly | 180 | 185 | 190 | |
| Trp | Leu | Asp | Trp | Leu | Lys | Met | Asp | Ile | Gly | Phe | Asp | Ala | Trp | Arg | Leu | 195 | 200 | 205 | |
| Asp | Phe | Ala | Lys | Gly | Tyr | Ser | Ala | Asp | Met | Ala | Lys | Ile | Tyr | Ile | Asp | 210 | 215 | 220 | |
| Ala | Thr | Glu | Pro | Ser | Phe | Ala | Val | Ala | Glu | Ile | Trp | Thr | Ser | Met | Ala | 225 | 230 | 235 | 240 |
| Asn | Gly | Gly | Asp | Gly | Lys | Pro | Asn | Tyr | Asp | Gln | Asn | Ala | His | Arg | Gln | 245 | 250 | 255 | |
| Glu | Leu | Val | Asn | Trp | Val | Asp | Arg | Val | Gly | Gly | Ala | Asn | Ser | Asn | Gly | 260 | 265 | 270 | |
| Thr | Ala | Phe | Asp | Phe | Thr | Thr | Lys | Gly | Ile | Leu | Asn | Val | Ala | Val | Glu | 275 | 280 | 285 | |
| Gly | Glu | Leu | Trp | Arg | Leu | Arg | Gly | Glu | Asp | Gly | Lys | Ala | Pro | Gly | Met | 290 | 295 | 300 | |
| Ile | Gly | Trp | Trp | Pro | Ala | Lys | Ala | Thr | Thr | Phe | Val | Asp | Asn | His | Asp | 305 | 310 | 315 | 320 |
| Thr | Gly | Ser | Thr | Gln | His | Leu | Trp | Pro | Phe | Pro | Ser | Asp | Lys | Val | Met | 325 | 330 | 335 | |
| Gln | Gly | Tyr | Ala | Tyr | Ile | Leu | Thr | His | Pro | Gly | Asn | Pro | Cys | Ile | Phe | 340 | 345 | 350 | |
| Tyr | Asp | His | Phe | Phe | Asp | Trp | Gly | Leu | Lys | Glu | Glu | Ile | Glu | Arg | Leu | 355 | 360 | 365 | |

Val Ser Ile Arg Asn Arg Gln Gly Ile His Pro Ala Ser Glu Leu Arg
 370 375 380

Ile Met Glu Ala Asp Ser Asp Leu Tyr Leu Ala Glu Ile Asp Gly Lys
 385 390 395 400

Val Ile Thr Lys Ile Gly Pro Arg Tyr Asp Val Glu His Leu Ile Pro
 405 410 415

Glu Gly Phe Gln Val Val Ala His Gly Asp Gly Tyr Ala Ile Trp Glu
 420 425 430

Lys Ile

- (2) INFORMATION FOR SEQ ID NO:7:
- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 709 base pairs
 (B) TYPE: nucleic acid
 (G) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA to mRNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
 (A) ORGANISM: Homo sapiens
- (vii) IMMEDIATE SOURCE:
 (B) CLONE: alpha-hemoglobin
- (ix) FEATURE:
 (A) NAME/KEY: transit_peptide (B) LOCATION:
 26. .241
 (B) LOCATION: 26. .241
- (ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 245. .670
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

CTCGAGGGCA TCTGATCTTT CAAGAATGGC ACAAATTAAC AACATGGCAC
 AAGGGATACA

60

AACCCCTTAAT CCCAATTCCA ATTTCCATAA ACCCCAAGTT CCTAAATCTT
 CAAGTTTCT

120

TGTTTTTGGG TGTAATAAAC TGAAAATTC AGCAAATCT ATGTTGGTTT TGAAAAAAGG

```

TTCAATTTTT ATGCAAAAGT TTTGTTTCCTT TAGGATTTCA GCAGGTGGTA
GAGTTTCTTG
240

CATG GTG CTG TCT CCT GCC GAC AAG ACC AAC GTC AAG GCC GCC TGG GGC
289

Val Leu Ser Pro Ala Asp Lys Thr Asn Val Lys Ala Ala Trp Cys
1 5 10 15
AAG GTT GGC GCG CAC GCT GGC GAG TAT GGT GCG GAG GCC CTG GAG AGG
Lys Val Gly Ala His Ala Gly Glu Tyr Gly Ala Glu Ala Leu Glu Arg
20 25 30
ATG TTC CTG TCC TTC CCC ACC ACC AAG ACC TAC TTC CCG CAC TTC GAG
Met Phe Leu Ser Phe Pro Thr Thr Lys Thr Tyr Phe Pro His Phe Asp
35 40 45
CTG AGC CAC GGC TCT GCC CAG GTT AAG GGC CAC GGC AAG AAG GTG GAG
Leu Ser His Gly Ser Ala Gln Val Lys Gly His Gly Lys Lys Val Ala
50 55 60
GAC GCG CTG ACC AAC GCC GTG GCG CAC GTG GAC GAC ATG CCC AAC GAG
Asp Ala Leu Thr Asn Ala Val Ala His Val Asp Asp Met Pro Asn Ala
65 70 75
CTG TCC GCC CTG AGC GAC CTG CAC GCG CAC AAG CTT CGG GTG GAC CAG
Leu Ser Ala Leu Ser Asp Leu His Ala His Lys Leu Arg Val Asp Pro
80 85 90 95
GTC AAC TTC AAG CTC CTA AGC CAC TGC CTG CTG GTG ACC CTG GCC GAG
Val Asn Phe Lys Leu Leu Ser His Cys Leu Leu Val Thr Leu Ala Ala
100 105 110
CAC CTC CCC GCC GAG TTC ACC CCT GCG GTG CAC GCC TCC CTG GAC AAG
His Leu Pro Ala Glu Phe Thr Pro Ala Val His Ala Ser Leu Asp Lys
115 120 125
TTC CTG GCT TCT GTG AGC ACC GTG CTG ACC TCC AAA TAC CGT
TAAGCTGGAG
677

Phe Leu Ala Ser Val Ser Thr Val Leu Thr Ser Lys Tyr Arg
130 135 140

CCTCGGTAGC CGTTCCTCCT GCCCGGTCGA CC

```

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 141 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(ix) SEQUENCE DESCRIPTION: SEQ ID NO:8:

| | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Val | Leu | Ser | Pro | Ala | Asp | Lys | Thr | Asn | Val | Lys | Ala | Ala | Trp | Gly | Lys |
| 1 | | | | 5 | | | | | 10 | | | | | 15 | |
| Val | Gly | Ala | His | Ala | Gly | Glu | Tyr | Gly | Ala | Glu | Ala | Leu | Glu | Arg | Met |
| | | | 20 | | | | | 25 | | | | | 30 | | |
| Phe | Leu | Ser | Phe | Pro | Thr | Thr | Lys | Thr | Tyr | Phe | Pro | His | Phe | Asp | Leu |
| | | 35 | | | | | 40 | | | | | 45 | | | |
| Ser | His | Gly | Ser | Ala | Gln | Val | Lys | Gly | His | Gly | Lys | Lys | Val | Ala | Asp |
| | 50 | | | | | 55 | | | | | 60 | | | | |
| Ala | Leu | Thr | Asn | Ala | Val | Ala | His | Val | Asp | Asp | Met | Pro | Asn | Ala | Leu |
| 65 | | | | | 70 | | | | | 75 | | | | | 80 |
| Ser | Ala | Leu | Ser | Asp | Leu | His | Ala | His | Lys | Leu | Arg | Val | Asp | Pro | Val |
| | | | | 85 | | | | | 90 | | | | | 95 | |
| Asn | Phe | Lys | Leu | Leu | Ser | His | Cys | Leu | Leu | Val | Thr | Leu | Ala | Ala | His |
| | | | 100 | | | | | 105 | | | | | 110 | | |
| Leu | Pro | Ala | Glu | Phe | Thr | Pro | Ala | Val | His | Ala | Ser | Leu | Asp | Lys | Phe |
| | | 115 | | | | | 120 | | | | | 125 | | | |
| Leu | Ala | Ser | Val | Ser | Thr | Val | Leu | Thr | Ser | Lys | Tyr | Arg | | | |
| 130 | | | | | | 135 | | | | | 140 | | | | |

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 743 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: Homo sapiens

(vii) IMMEDIATE SOURCE:

(B) CLONE: beta-hemoglobin

(ix) FEATURE:
(A) NAME/KEY: transit_peptide (B) LOCATION:
26..241
(B) LOCATION: 26..241

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 245..685

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

CTCGAGGGGA TCTGATCTTT CAAGAATGGC ACAAATTAAC AACATGGCAC
AAGGGATACA 60

AACCCTTAAT CCCAATTCCA ATTTCCATAA ACCCCAAGTT CCTAAATCTT
CAAGTTTTCT 120

TGTTTTTGGA TCTAAAAAAC TGAAAAATTC AGCAAATTCT ATGTTGGTTT
TGAAAAAAGA 180

TTCAATTTTT ATGCAAAAGT TTTGTTCTT TAGGATTTCA GCAGGTGGTA
GAGTTTCTTG 240

GATG GTG CAC CTG ACT CCT GAG GAG AAG TCT GCC GTT ACT GCC CTG TGG
289

Val His Leu Thr Pro Glu Glu Lys Ser Ala Val Thr Ala Leu Trp
1 5 10 15

GGC AAG GTG AAC GTG GAT GAA GTT GGT GGT GAG GCC CTG GGC AGG CTC 37

Gly Lys Val Asn Val Asp Glu Val Gly Gly Glu Ala Leu Gly Arg Leu
20 25 30

CTG GTG GTC TAC CCT TGG ACC CAG AGG TTC TTT GAG TCC TTT GGG GAG 55

Leu Val Val Tyr Pro Trp Thr Gln Arg Phe Phe Glu Ser Phe Gly Asp
35 40 45

CTG TCC ACT CCT GAT GCT GTT ATG GGC AAC CCT AAG GTG AAG GCT CAC 63

Leu Ser Thr Pro Asp Ala Val Met Gly Asn Pro Lys Val Lys Ala His
50 55 60

GGC AAG AAA GTG CTG GGT GCC TTT AGT GAT GGC CTG GCT CAC CTG GAG 71

Gly Lys Lys Val Leu Gly Ala Phe Ser Asp Gly Leu Ala His Leu Asp
65 70 75

AAC CTC AAG GGC ACC TTT GCC ACCA CTG AGT GAG CTG CAC TGT GAC AAG
529

Asn Leu Lys Gly Thr Phe Ala Thr Leu Ser Glu Leu His Cys Asp Lys
80 85 90 95

CTG CAC GTG GAT CCT GAG AGC TTC AGG CTC CTA GGC AAC GTG CTG ~~GTG~~ 7
 Leu His Val Asp Pro Glu Ser Phe Arg Leu Leu Gly Asn Val Leu Val
 100 105 110

TGT GTG CTG GCG CAT CAC TTT GGC AAA GAA TTC ACC CCA CCA GTG ~~CAG~~ 25
 Cys Val Leu Ala His His Phe Gly Lys Glu Phe Thr Pro Pro Val Gln
 115 120 125

GCT GCC TAT CAG AAA GTG GTG GCT GGT GTG GCT AAT GCC CTG GCC ~~CAG~~ 3
 Ala Ala Tyr Gln Lys Val Val Ala Gly Val Ala Asn Ala Leu Ala His
 130 135 140

AAG TAT CAC TAAGCTCGCT TTCTTGCTGT CCAATTTCTA TTAAAGGTTC 722
 Lys Tyr His
 145

CTTTGTGGGG TCGAGGTCGA C 743

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 146 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

Val His Leu Thr Pro Glu Glu Lys Ser Ala Val Thr Ala Leu Trp Gly
 1 5 10 15

Lys Val Asn Val Asp Glu Val Gly Gly Glu Ala Leu Gly Arg Leu Leu
 20 25 30

Val Val Tyr Pro Trp Thr Gln Arg Phe Phe Glu Ser Phe Gly Asp Leu
 35 40 45

Ser Thr Pro Asp Ala Val Met Gly Asn Pro Lys Val Lys Ala His Gly
 50 55 60

Lys Lys Val Leu Gly Ala Phe Ser Asp Gly Leu Ala His Leu Asp Asn
 65 70 75 80

Leu Lys Gly Thr Phe Ala Thr Leu Ser Glu Leu His Cys Asp Lys Leu
 85 90 95

His Val Asp Pro Glu Ser Phe Arg Leu Leu Gly Asn Val Leu Val Cys
 100 105 110

Val Leu Ala His His Phe Gly Lys Glu Phe Thr Pro Pro Val Gln Ala
115 120 125

Ala Tyr Gln Lys Val Val Ala Gly Val Ala Asn Ala Leu Ala His Lys
130 135 140

Tyr His
145

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: N-terminal

(vi) ORIGINAL SOURCE:
(A) ORGANISM: alkalophilic Bacillus sp.
(B) STRAIN: 38-2

(vii) IMMEDIATE SOURCE:
(B) CLONE: beta-cyclodextrin

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

Ala Pro Asp Thr Ser Val Ser Asn Lys Gln Asn Phe Ser Thr Asp Val
1 5 10 15

Ile